FACULTY OF LAW, ECONOMICS AND FINANCE

LAW RESEARCH PAPER SERIES

No. 2024 - 04

From B2B to B2G and G2G data sharing in competition law

When data become a competitive advantage and an enforcement tool

Author(s)

Isabella Lorenzoni
University of Luxembourg
isabella.lorenzoni@uni.lu

UNIVERSITY OF LUXEMBOURG

DL | DEPARTMENT OF LAW

From B2B to B2G and G2G data sharing in competition law. When data become a competitive advantage and an enforcement tool

Isabella Lorenzoni[*]

Abstract

Data has become an extremely important asset for our modern digital economy, mainly because of two key factors: the abundance of data that can be collected and processed (Big Data), and the technological advancement of Artificial Intelligence (AI) and data analytics techniques that make possible to transform data into meaningful information. In this scenario, data is not only a valuable resource for companies, but it is also a crucial element for public authorities when developing their own digital enforcement tools. This twofold function of data can be observed in competition law. Firstly, data serves as a competitive advantage for undertakings. By acquiring insights into users' habits and preferences, companies are able to tailor-made their products to customers' needs and strengthen their position in the market. Secondly, competition authorities have started to develop their own digital enforcement tools which would inevitably succeed or perish depending on data availability.

This paper analyses the crucial role that data plays in competition dynamics as a competitive advantage for undertakings and as an enforcement tool for competition authorities, addressing potential challenges. On the one hand, theories of harm that involve data collection have been developed, and remedies to ensure that data can be fairly accessed and shared are put forward. B2B data sharing can be achieved by relying firstly on competition law instruments that can offer *ex post* solutions, and secondly on regulations that can mitigate *ex ante* the risk of infringing competition law. On the other hand, competition authorities need to have access to huge amount of data (quantitative data) of the right kind (qualitative data) in machine readable format, in order to develop reliable and accurate digital enforcement tools. Therefore, mechanisms for encouraging data sharing and cooperation between businesses and competition authorities (B2G data sharing) and between competition authorities and other relevant public bodies (G2G data sharing) are here suggested.

Keywords

Competition Law, Data, Artificial Intelligence, Enforcement, Data sharing

[*] Isabella Lorenzoni is a doctoral researcher at the University of Luxembourg in Competition Law and Artificial Intelligence within the DILLAN program (Digitalisation Law and Innovation). Supported by the Luxembourg National Research Fund PRIDE 19/14268506. Email: isabella.lorenzoni@uni.lu.

1. Introduction

Every single action that we perform online produces data: from liking a post on social media to buying a bag on an online shop, and even by just clicking on a picture of a vacation island.[1] Data is nowadays considered the "lifeblood" of modern digital businesses.[2] Data is both, the input and the output of modern technologies on which digital platforms and businesses rely.[3] In such a scenario, it comes with no surprise that companies strive to collect personal and non-personal data, as it represents their main income source.[4]

This article aims to analyse the crucial role that data plays in competition dynamics as a competitive advantage for undertakings and as a potential enforcement tool for competition authorities. Firstly, solutions adopted to ensure data exchange between market players (B2B data sharing) are analysed, and secondly, recommendations to enhance cooperation and data sharing between private companies and competition authorities (B2G data sharing) and between competition authorities and other public enforcers (G2G data sharing) are laid down. In order to investigate the topic, section 2 is dedicated to analysing data as a resource for digital business models and its value for new technologies in the Artificial Intelligence (AI) chain. Section 3 delves into the topic of data as a competitive advantage and how competition law *ex post*, and regulations *ex ante* can mitigate the risk of abusing dominant positions and enhance B2B data sharing. Section 4 looks at the data from the side of enforcers, and how data has become a crucial element also for competition authorities to enforce competition law. Mechanisms for ensuring data sharing among different actors relevant for competition law are suggested, in the light of the current EU digital legislation which encourages data sharing between businesses and public authorities (B2G) and between public authorities among each other (G2G). Section 5 concludes.

[1] Lothar Determann and Teisha Johnson, 'Data Broker Regulation - Competition v. Privacy Considerations: Trade-Offs' (2023) CPI TechREG Chronicle <https://insightplus.bakermckenzie.com/bm/attachment_dw.action?attkey=FRbANEucS95NMLRN47z%2BeeOgEFCt8EGQJsWJiCH2WAWuU9AaVDeFgpCdzlkUxiWH&nav=FRbANEucS95NMLRN47z%2BeeOgEFCt8EGQbuwypnpZjc4%3D&attdocparam=pB7HEsg%2FZ312Bk8OIuOIH1c%2BY4beLEAeoUASUHJpPzQ%3D&fromContentView=1> accessed 16 September 2024.

[2] Marixenia Davilla, 'Is Big Data a Different Kind of Animal? The Treatment of Big Data Under the EU Competition Rules' (2017) 8 Journal of European Competition Law & Practice 370.

[3] OECD, 'Artificial intelligence, data and competition - Background Note' (2024).

[4] American Bar Association Antitrust Law Section, 'Artificial Intelligence & Machine Learning: Emerging Legal and Self-Regulatory Considerations, Part Two. Competition Implications of Big Data and Artificial Intelligence/Machine Learning' (Big Data Task Force, February 2021) <https://www.americanbar.org/news/abanews/aba-news-archives/2021/02/aba-antitrust-law-section-releases-part-two-of-white-paper-on-ai/> accessed 11 September 2024.

2. The importance of Data as input and output for modern business models

Data is often described as the "new oil of the 21st century",[5] and as the "new currency of the digital age".[6] Unsurprisingly, companies strive to acquire data, and just like unrefined oil, they need to process it, analyse it and transform it into valuable assets for their business.[7] Collecting and analysing data is not a practice that has emerged with the digital economy.[8] Companies have always been interested in acquiring users' data in order to learn their preferences and deliver products tailor-made to their needs. Marketing strategies revolve around this purpose: studying consumers' behaviours through collection and analysis of users' data.[9] What has changed now is the emergence of two essential factors: an enormous amount of data that can be harvested (Big Data) and technologies that allow to process raw data and extract valuable information (Big Data Analytics with AI and machine learning algorithms).[10] Digital companies, but also traditional brick and mortar stores are "avid collectors and users of data".[11] Collecting, processing and analysing data has become a successful strategy to compete for many business models.[12] One in particular, the multi-sided platform model, such as Google or Facebook, collects consumers' data from one side of the platform in exchange of services and products usually offered for free, and sell the data to the other side of the platform for targeted advertisements.[13]

---

[5] This expression was coined by the British mathematician Clive Humby in 2006 according to whom "Data is the new oil. It's valuable, but if unrefined it cannot really be used. It has to be changed into gas, plastic, chemicals, etc to create a valuable entity that drives profitable activity; so must data be broken down, analysed for it to have value." <https://towardsdatascience.com/is-data-really-the-new-oil-in-the-21st-century-17d014811b88> accessed 16 September 2024; <https://futurescot.com/why-data-is-the-new-oil/> accessed 16 September 2024; Sofia Ranchordás and Giovanni De Gregorio, 'Breaking Down Information Silos with Big Data: A Legal Analysis of Data Sharing' (2019) University of Groningen Faculty of Law Research Paper Series No. 44/2019 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3466313> accessed 16 September 2024; Davilla (n 2). See also the criticism expressed by Maggiolino and Colangelo to this expression: "Differently from an oil well, big data do not consist of a huge amount of the same product because, with the exception of duplicates, digital data are different from each other. […] although there is a market for oil, there is no market for a resource named "big data" and – even more importantly – there is no point in discussing whether firms compete in the market for big data or whether collectors, aggregators and analysers of big data operate in this same market. […] Neither can big data be compared with oil as such. It is well known that oil has some specific features: it is black and viscous. Big data do not have the same constant properties. Moving from one firm to another firm, not even the amount of data collected is the same". Giuseppe Colangelo and Mariateresa Maggiolino, 'Big Data as Misleading Facilities' (2018) Bocconi Legal Studies Research Paper No. 2978465, 3.

[6] Viktoria H S E Robertson, 'Excessive Data Collection: Privacy Considerations and Abuse of Dominance in the Era of Big Data' (2020) 57 Common Market Law Review 161. See also Maurice E. Stucke, 'Should we be concerned about Data-opolies?' (2018) 2 Georgetown Law Technology Review 275 and OECD, 'The intersection between competition and data privacy – Background Note' (2024) and Monika Woźniak-Cichuta, 'Digital Data-Driven Mergers: Is a Data-Sharing Remedy a Panacea?' (2024) 17 YARS 9.

[7] <https://towardsdatascience.com/is-data-really-the-new-oil-in-the-21st-century-17d014811b88> accessed 16 September 2024; <https://futurescot.com/why-data-is-the-new-oil/> accessed 16 September 2024. See also American Bar Association (n 4).

[8] Autorité de la Concurrence and Bundeskartellamt, 'Competition Law and Data' (2016).

[9] Ibid.

[10] Ibid.

[11] American Bar Association (n 4) 20.

[12] Ibid.

[13] Ibid. Autorité de la Concurrence and Bundeskartellamt (n 8).

Data can be classified in several ways. Big Data is typically described through the four "Vs": volume, velocity, variety, and value.[14] *Volume* refers to the amount of data collected which is a consequence of an increase in online activities by users with more interconnected devices, so-called Internet of Things (IoT), combined with lower costs for collecting, storing and processing data.[15] *Velocity* relates to the speed at which data are collected, processed and analysed, with nearly real-time data sharing.[16] *Variety* concerns the possibility of collecting multiple kinds of information from different sources, such as users' purchase history, websites visits, social media interactions.[17] All these features add *Value* to the data collected, which allow companies to generate revenues by making accurate predictions and targeting advertisements.[18]

Another classification of data refers to whether they are personal or non-personal data. According to the GDPR,[19] personal data refer to "any information relating to an identified or identifiable natural person (data subject)".[20] Non-personal data are any other data that are not personal which can be produced by interconnected devices (IoT, like sensors in self-drive cars) or they can derive from anonymisation of personal data, where the data subject is no longer identifiable.[21] Big Data usually comprises of both, personal and non-personal data whose separation in a given dataset can be rather difficult.[22]

Furthermore, data can be categorised depending on the ways in which they are obtained.[23] Firstly, data can be voluntarily or "actively" provided by individuals, who intentionally hand out their information, either because they want to do so (e.g. when filling consumers' forms with personal information) or because they do not have alternative choices (e.g. specific information required by a bank).[24] Secondly, data are passively provided by individuals, consciously or unconsciously, as their online activities are recorded (e.g. location tracking, web browsing) which are known under the name of "observed data".[25] Thirdly, data can be inferred through data analytics techniques which find correlations in the data previously collected (either as volunteered or observed data) to predict consumers behaviours and other characteristics (e.g. data inferred for credit scoring) which represent the real value for companies, as it is on the basis of inferred data that they can refine their

[14] Christophe Samuel Hutchinson, 'Potential abuses of dominance by big tech through their use of Big Data and AI' (2022) 10 Journal of Antitrust Enforcement 443; Davilla (n 2); Autorité de la Concurrence & Bundeskartellamt (n 8); OECD, 'Consumer Data Rights and Competition - Background note' (2020); OECD (n 3); Thomas Tombal, *Imposing Data Sharing among Private Actors A Tale of Evolving Balances* (Kluwer Law International 2022).
[15] Hutchinson (n 14); OECD (n 14); Davilla (n 2).
[16] Hutchinson (n 14); OECD (n 14); Davilla (n 2).
[17] Hutchinson (n 14); OECD (n 14); Davilla (n 2).
[18] Hutchinson (n 14); OECD (n 14); Davilla (n 2); Tombal (n 14).
[19] Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).
[20] Article 4(1) GDPR; Tombal (n 14); OECD (n 14) ; Autorité de la Concurrence and Bundeskartellamt (n 8).
[21] Tombal (n 14).
[22] Ibid.
[23] OECD (n 14); Tombal (n 14); Autorité de la Concurrence and Bundeskartellamt (n 8).
[24] Hutchinson (n 14); OECD (n 14); Tombal (n 14); Autorité de la Concurrence and Bundeskartellamt (n 8).
[25] Hutchinson (n 14); OECD (n 14); Tombal (n 14); Autorité de la Concurrence and Bundeskartellamt (n 8).

products or tailor advertisements.[26] Fourthly, data can also be acquired from third parties (e.g. from data brokers) through licensing agreements or through mandatory data sharing obligations.[27]

Finally, data are usually divided into unstructured and structured data.[28] When first collected, data are unstructured or "raw data" and they typically do not generate value unless transformed into structured data, from which meaningful information and knowledge can be extracted.[29] This process is nowadays possible thanks to Big Data analytics which employs machine learning techniques to draw inferences from data, which are in turn used for prediction or decision-making.[30]

Overall, data are valuable assets for companies that can dispose of AI and machine learning techniques to rapidly analyse and transform them into monetizable information.[31] At the same time, AI systems need to access large datasets to function, whose performance and level of accuracy increase when big data are fed into "data-hungry AI algorithms".[32] Hence data is both the input on which AI systems are trained and the output of AI applications, deployed to analyse data and extract relevant information.[33] Especially in large language models (LLMs), data is an extremely important requirement used to train foundation models. In fact, data plays a crucial role in each stage of the "GenAI value chain"[34]. Quantity and quality of data affect a foundation model's output, as "[a]n LLM will flourish or wither depending on the data it is trained on; and nothing can make up for that."[35]

---

[26] Hutchinson (n 14); OECD (n 14); Tombal (n 14); Autorité de la Concurrence and Bundeskartellamt (n 8).

[27] OECD (n 14); Tombal (n 14); Maureen K Ohlhausen, 'Privacy and Competition: Friends, Foes, or Frenemies?' (2019) CPI Antitrust Chronicle <https://www.competitionpolicyinternational.com/wp-content/uploads/2019/02/CPI-Ohlhausen.pdf> accessed 14 September 2024.

[28] Autorité de la Concurrence and Bundeskartellamt (n 8).

[29] Tombal (n 14); Autorité de la Concurrence and Bundeskartellamt (n 8); American Bar Association (n 4) "Data is typically not an end in itself, but rather is useful when inferences can be drawn from it" 9. See also Ranchordás and De Gregorio (n 5) "[…] 'raw data', that is, 'signs, patterns, characters, or symbols that can convey information about facts, concepts, processes, or objects.' Data becomes information when a 'datum' or 'given fact' is placed in a certain context and is hence labelled. 'Information' refers thus to the meaning assigned to data" 10. See also Colangelo and Maggiolino (n 5) according to whom "the value of big data lies in disclosing knowledge; that is, in revealing hidden pieces of information. Firms succeed in developing better products and services when they make data meaningful; that is, when they infer correlations and elaborate predictions as to consumers' tastes and rivals' strategies by using particular analytical tools" 23.

[30] Tombal (n 14); American Bar Association (n 4); Ranchordás and De Gregorio (n 5) "Data analytics employs for example algorithms to make sense of streams of data, identify relationships across unconnected datasets, and identify or predict behaviour and preferences" 6-7.

[31] Hutchinson (n 14).

[32] American Bar Association (n 4).

[33] OECD (n 3).

[34] "Foundation models are made by training a machine learning algorithm (called pre-training in this context) using huge datasets to produce a model that can be refined and used in many downstream applications. […] Fine-tuned models are foundation models that are refined through additional training on a narrower set of use case specific data. […] Grounded models have access to additional data sources, allowing the model access to information (e.g. real-time news) beyond the pre-training and fine-tuning data" Stefan Hunt, Wen Jian, Aman Mawar and Bartley Tablante, 'You Are What You Eat: Nurturing Data Markets to Sustain Healthy Generative AI' (2023) CPI 1 TechREG Chronicle <https://www.keystone.ai/wp-content/uploads/2023/11/NURTURING-DATA-MARKETS-TO-SUSTAIN-HEALTHY-GENAI-INNOVATION-Stefan-Hunt-Wen-Jian-Aman-Mawar-Bartley-Tablante-2.pdf> accessed 06 September 2024, 3.

[35] Hunt, Jian, Mawar and Tablante (n 34).

3. Data as a competitive advantage. B2B data sharing obligations

How companies collect, process and use consumers' data was usually seen as a matter for data protection or consumer law, and not a competition concern.[36] However, when data becomes a key factor around which companies compete and strengthen their market position, "the question of how that company handles the personal data of its users is not only relevant for data protection authorities, but also for competition authorities."[37] This section will first analyse different theories of harm that involve data and thereafter it will look at the mechanisms to impose *ex post* and *ex ante* data sharing obligations.

3.1 Data-driven theories of harm

Different theories of harm that involve the use of data have been conceived in competition law. Firstly, "*data-driven mergers*"[38] occur when, in order to increase data access, companies are incentivised to acquire other firms with the aim of combining and merging their datasets.[39] In data-driven markets, such mergers could provide a competitive advantage, as the new merging entity would have larger datasets and therefore more data to profile individuals.[40] In fact, "[t]he linkage of these data can give companies more insights into user habits, enabling them to further improve their services and reinforce their market position. Generally speaking, the more data a company can combine, the better its chances to gain knowledge that can be used to strengthen its market position."[41] Concentration of data could raise competition concerns, as the merging firm may enhance its market power and raise barriers to entry and costs for competitors who may not be able to replicate the data resulting from combination of their datasets.[42] In the past, several data-driven

[36] Autorité de la Concurrence and Bundeskartellamt (n 8); OECD (n 6).

[37] Bundeskartellamt, 'Bundeskartellamt prohibits Facebook from combining user data from different sources Background information on the Bundeskartellamt's Facebook proceeding' (2019) 6 <https://www.bundeskartellamt.de/SharedDocs/Publikation/EN/Diskussions_Hintergrundpapiere/2019/07_02_2021 9_Hintergrundpapier_Facebook.pdf?__blob=publicationFile&v=2> accessed 05 August 2024. See also OECD (2024) (n 6); Peter Stauber, 'Facebook's Abuse Investigation in Germany and Some Thoughts on Cooperation Between Antitrust and Data Protection Authorities' (2019) 2 CPI Antitrust Chronicle 36 according to whom "[i]f access to data is an essential factor for the competitive position of the company – as is the case with data-driven products such as social networks – the handling of personal data by a company is not only a case for data protection authorities, but also for antitrust authorities" 6.

[38] Christophe Carugati, 'The Antitrust Privacy Dilemma' (2021) working paper <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3968829> accessed 28 August 2024. "The notion of data-driven mergers refers to transactions which are motivated by an acquisition of a huge amount of data" in Woźniak-Cichuta (n 6) 14.

[39] Autorité de la Concurrence and Bundeskartellamt (n 8), OECD (n 6) and Woźniak-Cichuta (n 6).

[40] Giuseppe Colangelo and Mariateresa Maggiolino, 'Data Protection in Attention Markets: Protecting Privacy through Competition?' (2017) 8 Journal of European Competition Law & Practice 363; Autorité de la Concurrence and Bundeskartellamt (n 8) and Woźniak-Cichuta (n 6).

[41] Nils-Peter Schepp and Achim Wambach, 'On Big Data and Its Relevance for Market Power Assessment' (2016) 7 Journal of European Competition Law & Practice 120, 121 and Daniele Condorelli and Jorge Padilla, 'Harnessing Platform Envelopment in the Digital World' (2020) 16 Journal of Competition Law & Economics 143. See also Woźniak-Cichuta (n 6), according to whom "the value comes from the large combination of a variety of data collected from different sources (e.g., data analytics' sophistication), […] the incentive for big data firms will be to aim to expand the variety of data collected" 36.

[42] Autorité de la Concurrence and Bundeskartellamt (n 8); OECD (n 6); OECD (n 14); Davilla (n 2).

mergers have been cleared at the EU level, despite concerns that combination of datasets would have given rise to anticompetitive issues.[43]

Secondly, when it comes to *collusive practices and cartels* under Article 101 Treaty on the Functioning of the European Union (TFEU), agreements between companies aimed at maximising data collection and consumer information may be seen as a competition concern.[44] Collusion may also occur when data are illegally shared among competitors to get insights of respective commercial activities and competitive strategies, with the aim of coordinating their actions instead of competing against each other.[45]

Thirdly, in data-driven markets, abuse of dominance, according to Article 102 TFEU, can occur under two main theories of harm: *exclusionary abuses* and *exploitative abuses*. Exclusionary conducts refer to practices that have a foreclosure effect by denying access to other competitors.[46] This can happen when a company who has a dominant position refuses to deal and to provide access to a fundamental input without which rivals cannot compete in a market.[47] Refusal to access data may constitute an abuse of dominant position if data are considered an "essential facility".[48] Furthermore, as an exclusionary conduct, "where a dominant firm has exclusive access to consumer data, it could attempt to raise rivals' costs or barriers to entry by engaging in tying or bundling."[49] Finally exploitative abuses can take the form of excessive data collection of users' data, when a company reduces the level of privacy protection and increase collection of data which is excessive or unfair.[50] The leading case is *Meta Platforms*,[51] where the German Competition authority (the Bundeskartellamt) found Facebook infringing competition law for abusing its dominant position in the market for social networks, by imposing unfair terms and conditions that resulted in excessive collection of users' data.[52]

3.2 Data as an essential facility. *Ex post* B2B data sharing

[43] See for instance *Google/DoubleClick* (Case COMP/M.4731) Commission decision [2008] OJ C184/10; *Facebook/WhatsApp* (Case COMP/M.7217) Commission Decision [2014] OJ C417/4; *Microsoft/LinkedIn* (Case COMP/M.8124) Commission decision [2016] OJ C 388/4. See also OECD (n 6); OECD (n 14); Davilla (n 2); Colangelo and Maggiolino (n 40) and Woźniak-Cichuta (n 6).
[44] OECD (n 14).
[45] Davilla (n 2). See also Woźniak-Cichuta (n 6) "[a]nother challenging aspect of data-sharing remedies is that they can facilitate the creation of anticompetitive agreements" 37. For a deeper analysis on the risk of collusion brought by data sharing see Eugenio Olmedo-Peralta, 'The Creation of Data Pools as Information Exchanges: Antitrust Concerns' (2024) 17 YARS 49 at 70 ss.
[46] Hutchinson (n 14).
[47] Ibid.
[48] Ibid; Autorité de la Concurrence and Bundeskartellamt (n 8); OECD (n 6); OECD (n 14) and Davilla (n 2).
[49] OECD (n 14). See also Hutchinson (n 14) and Autorité de la Concurrence and Bundeskartellamt (n 8), according to whom tying practices can happen when "a company owning a valuable dataset ties access to it to the use of its own data analytics services. As it noted, such tied sales may increase efficiency in some circumstances but they could also reduce competition by giving a favorable position to that company which owned the dataset over its competitors on the market for data analytics" 20.
[50] Stucke (n 6); OECD (n 6); OECD (n 14); Hutchinson (n 14), Davilla (n 2).
[51] See Decision B6-22/16 of 6 February 2019 of the Bundeskartellamt and C-252/21 *Meta Platforms Inc. et al. v Bundeskartellamt* [2023] EU:C:2023:537.
[52] See for instance Stauber (n 37); OECD (n 6); OECD (n 14); Hutchinson (n 14).

Data is considered a "non-rivalrous" good, which means that the use of certain data by a company does not prevent, in principle, other companies from accessing and using the same data.[53] From a competition point of view, data can enhance innovation and maintain competition on the markets. However, data may also represent a competitive advantage for few big companies if certain data cannot be replicated by competitors.[54] For example, looking at large language models, which are trained on billions of data, even though many datasets are publicly available and open source, "some large datasets are proprietary and may provide unique insights that others struggle to replicate, noting as an example Google's ownership of YouTube and the potential to control access to its video transcripts".[55] Hence, specific data could be a crucial factor for companies to compete, and therefore they may be regarded as "close to essential services".[56]

If data can be considered an essential input without which rivals would not be able to compete, competition law may impose compulsory B2B data sharing as an *ex post* remedy, when refusal to share an essential resource (refusal to deal) amounts to an abuse of dominant position under Article 102 TFEU.[57] The tool traditionally used to restore competition in such cases is the "Essential Facility Doctrine (EFD)" and only in exceptional circumstances access can be obtained.[58]

The EFD is considered "one of the most controversial antitrust issues",[59] as it imposes to a monopolist that owns an essential input an obligation to share it with anyone who needs it, including competitors, limiting its rights to property and its freedom to contract.[60] When granting access to a crucial facility, it is important to strike a balance between the risk of dominant undertakings to abuse their position by excluding others from the facility they own, extending their

[53] Davilla (n 2); American Bar Association (n 4) 13 "[d]ata is typically non-rivalrous. At the outset a firm's use of most data does not typically reduce the data's availability to competitors—that is, data is generally non-rivalrous. Data can be replicated, provided to multiple suppliers by buyers, shared between businesses, and gathered by multiple entities, all of which make market power through data less likely". See also Colangelo and Maggiolino (n 5) 6.
[54] OECD (n 3) 30.
[55] OECD (n 3). See also Hunt, Jian, Mawar and Tablante (n 34) according to whom "superior access to data could give some players a significant advantage: the largest AI players have access to proprietary datasets, e.g. YouTube data by Google, and train their models on them" 24-25.
[56] OECD (n 3) 30. For example, in the field of Generative AI, Hunt, Jian, Mawar and Tablante (n 34) suggest that "[g]iven the potential for such challenges, markets for data for pre-training, tuning, and grounding might need nurturing from regulators, to preserve healthy GenAI competition and alleviate consumer protection issues. Agencies might need to start getting "under the hood" more actively, for instance by monitoring the data AI companies are accessing and using. Regulators should also consider making this information at least partially available to some parties through transparency requirements, as the EU is proposing with the AI Act. Other regulatory responses to nurture data markets could include monitoring for harmful exclusionary vertical agreements, requirements on data sharing or other rules" 25.
[57] Tombal (n 14); Oscar Borgogno and Giuseppe Colangelo, 'Data sharing and interoperability: Fostering innovation and competition through APIs' (2019) 35 Computer Law & Security Review 1 and Hutchinson (n 14).
[58] Borgogno and Colangelo (n 57); Colangelo and Maggiolino (n 5). The *hiQ v LinkedIn* case provides an interesting example of data as an essential facility *hiQ Labs, Inc. v. LinkedIn Corporation*, No. 17-cv-03301-EMC. See Ohlhausen (n 27) and Tombal (n 14).
[59] Colangelo and Maggiolino (n 5) 13 and Borgogno and Colangelo (n 57).
[60] Colangelo and Maggiolino (n 5); Borgogno and Colangelo (n 57) and Tombal (n 14). See also C-48/22 P *Google LLC, Alphabet Inc. (Google Shopping) v Commission* [2024] EU:C:2024:726, [91]; opinion of Advocate General Kokott C-48/22 P *Google LLC, Alphabet Inc. (Google Shopping) v Commission* [2024] EU:C:2024:14, [86] and opinion of Advocate General Medina C-233/23 *Alphabet Inc. et al. v Autorità Garante della Concorrenza e del Mercato* [2024] EU:C:2024:694, [32].

monopoly to other markets, and the risk of competitors to free ride on investments made by dominant companies to obtain the essential input.[61]

According to the EFD, as construed by the European Court of Justice in several case law, refusal by a dominant undertaking to provide access to its own facility to other undertakings may be an abusive conduct only in exceptional circumstances and when some conditions are met.[62] Firstly, the resource for which access is sought must be indispensable for carrying out a business activity on a secondary market. As specified in *Bronner*,[63] in order to be indispensable, no actual or potential alternatives must be available, and it is not enough to argue that it is not economically viable.[64] There must be technical, legal or economic obstacles or being unreasonably difficult for any undertaking to replicate the essential facility, alone or in cooperation with other companies.[65] Secondly, such refusal should likely exclude competition on a secondary market that the dominant company reserves to itself, and according to the *IMS Health* case,[66] it is sufficient to identify a potential or hypothetical market.[67] Thirdly, the refusal should prevent the appearance of a new product or service, or of technical improvements (*Microsoft* case[68]), for which there is potential consumer demand and no offer by the dominant undertaking.[69] Fourthly, there should be no objective justifications for such refusal.[70]

Comparing to traditional resources, data present specific characteristics, and some scholars claim that the EFD should be adapted accordingly when applied to a refusal to share data.[71] As far as the first condition is concerned, access seekers must still demonstrate the indispensability of data, as raw or unstructured data, when such data cannot be replicated and it is not possible to find a

[61] Colangelo and Maggiolino (n 5) and Borgogno and Colangelo (n 57). Tombal (n 14) "a balance must be found between incentivising innovation (data collection and processing by data holders) and maximising social welfare through large dissemination (data sharing)" [224]. See also AG Kokott opinion C-48/22 P (n 60) [87].
[62] Tombal (n 14). On the conditions to apply the EFD see more recently C-48/22 P *Google Shopping* (n 60) [89, 90, 109 – 112]. See also AG Medina opinion C-233/23 (n 60) according to whom "in order to determine whether the Bronner conditions apply to a case concerning a refusal to grant access, it is necessary to discern whether the infrastructure to which access is requested is to be consecrated to the dominant undertaking's own business and use and to be enjoyed exclusively by it, as a way of preserving the benefits of the investments made for the development of that infrastructure. By contrast, those conditions are not intended to apply where the infrastructure concerned is opened to other operators on the market" [35]. See also [28, 56].
[63] C-7/97 *Bronner* [1998] EU:C:1998:569.
[64] *Bronner* [45].
[65] *Bronner* [44]. Tombal (n 14); Colangelo and Maggiolino (n 5).
[66] C-418/01 *IMS Health* [2004] EU:C:2004:257.
[67] *IMS Health.* Tombal (n 14); Colangelo and Maggiolino (n 5) "It is sufficient to demonstrate the existence of two interconnected markets rather than proving the effective commercialization of the essential inputs" 17.
[68] T-201/04 *Microsoft v. Commission* [2007] EU:T:2007:289.
[69] Tombal (n 14); Colangelo and Maggiolino (n 5).
[70] Tombal (n 14); Colangelo and Maggiolino (n 5).
[71] For instance, according to Bertin Martens, Alexandre de Streel, Inge Graef, Thomas Tombal and Néstor Duch-Brown, 'Business-to-Business data sharing: An economic and legal analysis' (2020) Digital Economy Working Paper 2020-05, European Commission, JRC121366 at 36-37 "[…] the non-rivalry and wide availability of data will often render a dataset non indispensable" and therefore when referring to access to data "an expansion of the interpretation of the essential facilities doctrine beyond cases of leveraging and a lower threshold for the condition of indispensability (as well as new product)" should apply. See also Tombal (n 14) and Davilla (n 2).

substitute.[72] Only in exceptional circumstances, when the "data holder is the sole source of a specific type of data, the access to which is indispensable for another undertaking evolving on a downstream market",[73] the EFD may apply as an *ex post* compulsory B2B data sharing remedy.[74] Furthermore, according to Tombal, the threshold for the second and third conditions should be lowered.[75] Finally, the last condition of objective justifications must be thoroughly assessed by competition authorities, especially when big tech companies refuse to share data on the ground of protection of users' data, which may only be a defence shield for abusing their dominant position.[76]

Other scholars are of different opinions. For instance, Davilla suggests that instead of relying on the EFD, whose conditions may be too difficult to apply when data are involved, an *ad hoc* test which takes into consideration the specificities of data should be developed.[77] Colangelo and Maggiolino strongly oppose the application of the EFD to data for several reasons.[78] Firstly, they state that the third condition which requires a new or improved product may be unsuitable for data, as access seekers usually do not know in advance what type of product they may develop with the data obtained as they may not know the information that data can reveal.[79] Arguably, data "are not gathered and organized in order to answer specific research questions. Rather the opposite: firstly, they are amassed from many and varied sources, then they are analysed to make them speak".[80] Secondly, according to them, data is the "wrong target", as what is truly essential is not the data as such but the information they provide and the analytical tools used to extract such information, as "just by looking at big data no one can guess the information they include and – a fortiori – no one can say in advance what pieces of information among the many that can be extracted from big data will be essential […]".[81]

When imposing data sharing as an *ex post* remedy to restore competition on the market, attention should be paid to the categories of data that the data holder should share with the undertaking

[72] Tombal (n 14); Autorité de la Concurrence and Bundeskartellamt (n 8) such requirement "would only be met, if it is demonstrated that the data owned by the incumbent is truly unique and that there is no possibility for the competitor to obtain the data that it needs to perform its services" 18.
[73] Tombal (n 14) [241].
[74] Tombal (n 14).
[75] Ibid.
[76] "[C]ompetition authorities should require these large data holders to lay down and substantiate the data protection concerns they raise in order to refuse data sharing with third parties, and they should collaborate with data protection authorities in order to assess whether the data protection standards imposed on third parties by these large firms are (suspiciously) higher than the ones they apply to themselves" Thomas Tombal, 'Data protection and competition law: friends or foes regarding data sharing?' (Accepted paper for the TILTing Perspectives 2021 Conference: Regulating in Times of Crisis) 22. See also Tombal (n 14) 255; OECD (n 6) and Woźniak-Cichuta (n 6).
[77] Davilla (n 2) 380. See also Woźniak-Cichuta (n 6), according to whom "[…] in order to formulate a data-sharing remedy […] other criteria should be formulated, rather than those set out in the essential facilities doctrine" 40; and Olmedo-Peralta (n 45) according to whom "it is fair to view the application of the essential facilities doctrine to data pools as not appropriate, considering the inherently duplicable nature of data, the multiple ways in which they can be accessed, and the possibility of building competing data consortia, as no one controls an indispensable non-duplicable resource for the market. In this case, it will be a market failure that must be approached through other regulatory instruments, theories of harm, or applicative tools" 69.
[78] Colangelo and Maggiolino (n 5).
[79] Ibid; and Giuseppe Colangelo and Mariateresa Maggiolino, 'Data access and AI: Antitrust vs. Regulation' (2018).
[80] Colangelo and Maggiolino (n 5) 21.
[81] Colangelo and Maggiolino (n 5) 24.

which seeks access to it.[82] Firstly, data that are commercially available in data marketplaces should not be subject to the remedy, as they can be acquired from data brokers or other third parties, just like the dominant undertaking.[83] Secondly, according to the previous distinction, only volunteered and observed data should fall within the scope of a data sharing remedy as "observed data, […] often cannot be replicated, and volunteered data […] would take a significant amount of effort to volunteer again".[84] Such volunteered and observed data need to be indispensable for the undertaking that seeks access to it, in order to carry out its business in a secondary market.[85] Finally, inferred data should fall out the scope of the remedy, because they are the result of other investments made by the data holder that involve the use of Big Data analytics in the form of machine learning and AI-applied solutions to extract valuable information for the company to innovate.[86] Hence, once raw data (volunteered and observed data) are shared with undertakings requesting such access, it is for them to apply their own technology and extract relevant insights for their business.[87] However, the remedy should only cover a subset of data and not all the data of the dominant undertaking[88] which may be a difficult task, given the fact that big data are usually collected "untidily".[89] Another problem when imposing B2B data sharing is to consider whether one time data access would be able to restore competition, as according to the life cycle of data, access may be needed "on a constant basis and in (near) real time".[90] However, imposing long-term data access commitments may be an excessively restrictive measure which would result in over-enforcement.[91]

Finally, data protection issues should be carefully assessed when imposing B2B data sharing as an *ex post* remedy. For instance, in the *GDF Suez* case,[92] the French competition authority obliged the company to share its customers' list with other competitors, and the data protection authority expressed concerns about infringing data protection rules as personal data were contained in the list. Therefore, users were given the choice to object to share their data with other companies.[93] This calls for a more coordinated approach and cooperation between competition and data protection authorities, whenever *ex post* compulsory data sharing may raise data protection concerns.[94] Furthermore, data protection issues could be overcome by implementing

[82] Tombal (n 14).
[83] Ibid; Colangelo and Maggiolino (n 5).
[84] Jacques Crémer, Yves-Alexandre de Montjoye and Heike Schweitzer, 'Competition Policy for the digital era' (2019) Final report for the European Commission <https://op.europa.eu/en/publication-detail/-/publication/21dc175c-7b76-11e9-9f05-01aa75ed71a1/language-en> accessed 16 September 2024, 101. See also Tombal (n 14).
[85] Tombal (n 14).
[86] Ibid; Crémer, de Montjoye and Schweitzer (n 84).
[87] Tombal (n 14).
[88] Tombal (n 14); Colangelo and Maggiolino (n 5) "some digital data".
[89] Colangelo and Maggiolino (n 5).
[90] Tombal (n 14) [305]. See also Colangelo and Maggiolino (n 5).
[91] Woźniak-Cichuta (n 6).
[92] <https://www.autoritedelaconcurrence.fr/en/communiques-de-presse/9-september-2014-gas-market> accessed 16 September 2024.
[93] OECD (n 6); Carugati (n 38); Christophe Carugati, 'Overview of privacy in cases relevant to competition law' [2023] Concurrences 1.
[94] OECD (n 6); Carugati (n 38).

anonymisation[95] and pseudonymization[96] techniques, especially when access to data is necessary for the purpose of developing AI tools, as "AI may need a huge and diversified amount of data, but it does not necessarily want to understand who individuals are and what they do: not necessarily AI needs data that serve to identify single persons or that make them identifiable. AI can work also with anonymized and pseudonymized data".[97]

3.3 *Ex ante* B2B data sharing

Compulsory B2B data sharing can also be imposed by regulations as an *ex ante* remedy, before starting an investigation and before causing harm to consumers and competitors,[98] to remedy to situations where refusal to share data may amount to anticompetitive behaviours under specific circumstances.[99] Firstly, *ex ante* B2B data sharing can be imposed by sectorial legislations[100] which have the advantage of being "more targeted and adapted to the sector's needs, characteristics and data standardisation challenges".[101] However, sectorial legislations may present as a downside to provide a data advantage to large data holders, who could use sectorial data to refine other data-driven services that they offer, strengthening their position in those markets, as "these large data holders [would] have a 360° view on the consumers' preference".[102]

Secondly, compulsory B2B data sharing can also be imposed by horizontal regulations when specific circumstances require an *ex ante* intervention to remedy to market failures, such as data concentration and data conglomeration, that cannot be dealt efficiently only by relying on *ex post* competition enforcement law.[103] Accordingly, "[…] the antitrust enforcement toolbox is inadequate to tackle effectively the need to ensure access to data. The scope of competition law is limited by the fact that it can be invoked only to gain access to a dataset held by a dominant firm,

---

[95] "Anonymization refers to the process of either encrypting or removing personally identifiable information from datasets, such that the people whom the data describe (data subjects) remain anonymous" Jens Prüfer, 'Competition Policy and Data Sharing on Data-driven Markets Steps Towards Legal Implementation' (Friedrich-Ebert-Stiftung project 2020) <https://library.fes.de/pdf-files/fes/15999.pdf> accessed 26 August 2024, 12.

[96] "Pseudonymization refers to the process of replacing personally identifiable information fields by one or more artificial identifiers, or pseudonyms" Prüfer (n 95) 12.

[97] Colangelo and Maggiolino (n 79) 6.

[98] Tombal (n 14) and OECD, 'G7 inventory of new rules for digital markets: Analytical note' (2023).

[99] OECD (n 98).

[100] See for instance in the sector of banking with the PSD2 Directive (Directive (EU) 2015/2366 of 25 November 2015 on payment services in the internal market, amending Directives 2002/65/EC, 2009/110/EC and 2013/36/EU and Regulation (EU) No 1093/2010, and repealing Directive 2007/64/EC) and in the automotive sector (Regulation (EU) 2018/858 of 30 May 2018 on the approval and market surveillance of motor vehicles and their trailers, and of systems, components and separate technical units intended for such vehicles, amending Regulations (EC) No 715/2007 and (EC) No 595/2009 and repealing Directive 2007/46/EC), which imposes on car manufacturers to share data of on-board diagnostics, repair and maintenance with independent operators in machine readable format. Tombal (n 14) [373] and Martens *et al.* (n 71).

[101] Tombal (n 14) [372].

[102] Ibid [375].

[103] Tombal (n 14); Giuseppe Colangelo, 'The European Digital Markets Act and antitrust enforcement: a liaison dangereuse' (ICLE White Paper 2022); OECD, 'Ex ante regulation of digital markets' (2021) OECD Competition Committee Discussion Paper; Daniel Pettersson, 'Sector-Specific Ex Ante Regulation in Digital Markets – A Complement or Substitute to Antitrust Enforcement?' (2022) 4 Europarättslig tidskrift. For instance, Recital 5 DMA states that enforcement of Articles 101 and 102 TFEU "[…] occurs ex post and requires an extensive investigation of often very complex facts on a case by case basis".

on a case-by-case basis".[104] Instead, *ex ante* regulations can "impose clear upfront before-the-event obligations that will be of more straightforward and speedy application compared to ex post enforcement, which assesses the conduct and its illegality once it has occurred".[105] For instance, when it comes to compulsory data sharing, the DMA[106] provides rules on mandatory data access, according to which gatekeepers shall provide business users with access to the data generated by their activities and interactions on the platform.[107]

Ex ante regulations may impose obligations on specific and well-identified data holders[108] to make some of their datasets accessible to certain business users.[109] As it was the case for *ex post* data sharing remedies, *ex ante* regulations should only require to share volunteered and observed data, leaving inferred data outside their scope, as they are more valuable for the data holder.[110] Furthermore, contrary to *ex post* competition law remedies, *ex ante* regulations can impose obligations on data receivers who could be required to only use the data obtained for a specific legitimate and economic purpose that has been declared in advance and to provide adequate guarantees for privacy and data protection, preventing reidentification in case the data shared have been previously pseudonymised or anonymised.[111]

As already seen with *ex post* B2B data sharing obligations, privacy and data protection rules may create some frictions with *ex ante* B2B data sharing regulations. Also in this case, anonymization and pseudonymization should be encouraged as techniques to safeguard individuals' personal data, carefully considering costs and benefits for both, data holders and data recipients.[112] Furthermore, in compliance with the GDPR, *ex ante* rules may provide a legal basis to impose data sharing obligations to the controller,[113] if such an obligation is "clear and precise and its application [is] foreseeable to persons subject to it".[114]

From a technical point of view, *ex ante* B2B data sharing rules can be successfully implemented only if specific tools are available that allow data transition between companies in machine readable format.[115] For this reason, Application Programming Interfaces (APIs)[116] with the

[104] Borgogno and Colangelo (n 57).
[105] OECD (n 98) 5.
[106] Regulation (EU) 2022/1925 of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act).
[107] Article 6(10) DMA. A similar provision can be found for example in Section 3(a)(7) of the American Innovation and Choice Online Act <https://www.congress.gov/bill/117th-congress/senate-bill/2992/text> accessed 03 September 2024. See OECD (n 98) and OECD, 'G7 inventory of new rules for digital markets. Prepared by the OECD Competition Division for the 2023 G7 Joint Competition Policy Makers and Enforcers Summit' (2023).
[108] Tombal (n 14) "[…] namely those that are considered as playing a central role in the (systemic) market failure(s) that led to the legislative intervention" [385].
[109] Tombal (n 14).
[110] Ibid.
[111] Ibid.
[112] Ibid.
[113] Article 6 and Recital 111 GDPR.
[114] Recital 41 GDPR. See also Tombal (n 14).
[115] Tombal (n 14); Borgogno and Colangelo (n 57) and Woźniak-Cichuta (n 6).
[116] "APIs can be defined in broad terms as software tools designed to enable communication between two computer applications. Through a set of protocols and routines, they allow a digital application to interact with an associated program by describing the kind of data that can be retrieved, how to do it and the format in which information will be filed" Borgogno and Colangelo (n 57).

adoption of common technical standards and protocols are essential to enable interoperability[117] between the systems and ensure an efficient implementation of data sharing rules.[118] However, imposing specific technical standards may add substantial costs for both, data holders and data recipients.[119] Hence, instead of leaving data holders and data recipients free to organize how data sharing should occur, another solution that has been suggested is to establish intermediary bodies to act as "data trustees" entrusted with the task of facilitating data exchange between the parties, where the data holder would not share data with the recipients but only with the intermediary body.[120] Furthermore, algorithms could be trained on the intermediaries' dataset, which contains personal and non-personal data, but instead of having a human being accessing those data, only trained algorithms – and not the original data shared by the data holder – would be transferred to the data recipients.[121]

Overall, both *ex post* and *ex ante* B2B data sharing obligations can provide suitable solutions to remedy the problem of data as a competitive advantage in specific circumstances. By relying on the EFD, competition law provides an *ex post* remedy, by imposing an obligation on the data holder to share some of their data with the access seeker, only when exceptional circumstances are met, which is necessarily based on a "case-by-case approach".[122] On the contrary, regulations may impose compulsory B2B data sharing as an *ex ante* remedy, which can offer "universal and general solutions, which may address whole industries, sectors, and market"[123] but at the same time be less tailor-made for specific needs.

4. Data as a competition enforcement tool. B2G and G2G data sharing

The last section of this paper looks at data from a different perspective, that is how data can help authorities to enforce competition law. Besides being an extremely valued resource for many business models, data can be a game changer also for competition authorities. In fact, competition authorities require access to data in order to develop their own digital enforcement tools to enhance their detection powers and strengthen their proactive approach.[124] Any digital tool based on AI and machine learning systems that competition authorities would like to internally develop require qualitative and quantitative data (e.g. digital screening tools used in the initiation phase to flag markets' anomalies and to start *ex officio* investigations).[125] "The challenges competition authorities face are the existence of data, to begin with, access to such data (and in particular

[117] "Interoperability refers to the ability of products and services at different levels (vertical interoperability) or at the same level (horizontal interoperability) of the digital value chain to work together" OECD (n 6) 21. See also OECD (n 98); Crémer, de Montjoye and Schweitzer (n 84); Borgogno and Colangelo (n 57).
[118] Tombal (n 14); Borgogno and Colangelo (n 57) according to whom "APIs' architecture and design has been identified as a crucial element for a flourishing common European data space" 5.
[119] Tombal (n 14).
[120] Prüfer (n 95); Tombal (n 14). See also Olmedo-Peralta (n 45) concerning data pool contracts and the use of a third-party intermediary.
[121] Prüfer (n 95) and Tombal (n 14).
[122] Borgogno and Colangelo (n 57); Colangelo and Maggiolino (n 79) and Martens *et al.* (n 71).
[123] Colangelo and Maggiolino (n 79).
[124] OECD, 'Data Screening Tools in Competition Investigations' (2022) OECD Competition Policy Roundtable Background Note.
[125] Ibid; Herwig C H Hofmann and Isabella Lorenzoni, 'Future Challenges for Automation in Competition Law Enforcement' (2023) 3 Stanford Computational Antitrust 36.

disaggregated and raw data, and data that are not publicly available), format, integrity and quality of data, and data searchability, cleaning and use."[126] Competition authorities need to access "relevant, up-to-date and structured business data"[127] in order to develop their own digital enforcement tools. Hence, data represents the key element also for competition authorities to move towards a "computational antitrust"[128] and to embrace the digital antitrust revolution. Data is the necessary glue that tights together technological innovations used for enforcement purpose.

Firstly, competition authorities would require access to *qualitative data*, in order to avoid inaccurate and useless results.[129] Secondly, data need to be abundant (*quantitative data*) in order to create a large database on which algorithms can be trained. For instance, digital screening tools used to detect bid rigging cases in public procurement require a large dataset containing enough examples of past cases of collusive and non-collusive conducts.[130] Thirdly, data need to be already in *machine readable format* to avoid time consuming manual conversion process.[131]

How can competition authorities have access to such data – qualitative and quantitative data in machine readable format? Arguably, they can first rely on publicly available information that can be found in companies' registers, and chambers of commerce[132] (even though it is unlikely that all data will be in the right format, but further treatment may be required). Second, web scraping techniques could be used to extract relevant information from websites.[133] Third, competition authorities, just like any other public and private actor, can purchase data from data brokers and third-party providers.[134] However, not all data that competition authorities may need are publicly available or can be purchased from third data providers, nor web scraping techniques may provide a long-term solution as it is considered a time-consuming mechanism.[135] Therefore, further solutions should be researched to enable competition authorities to access relevant data that can be used as an enforcement tool to feed their own digital systems. How can data access for competition authorities be improved? Two different channels are here proposed: imposing B2G data sharing and strengthening G2G cooperation and data sharing.

4.1. Imposing B2G data sharing

The first solution proposed is to impose B2G data sharing which would enable competition authorities to get access to privately held data outside their investigation powers.[136] Imposing

[126] OECD (n 124).
[127] Viktoria H S E Robertson and Jürgen Fleiß, 'Computational Antitrust and the Future of Competition Law Enforcement' [2024] GRUR International XX(XX) 1.
[128] Thibault Schrepel, 'Computational Antitrust: An Introduction and Research Agenda' (2021) 1 Stanford Computational Antitrust 1.
[129] OECD (n 124).
[130] Ibid and Ioannis Lianos, 'Computational Competition Law and Economics: Issues, Prospects – An Inception Report' (2021) Hellenic Competition Commission
[131] OECD (n 124).
[132] Ibid. See also OECD, 'Detecting cartels for ex officio investigations' (2024) OECD Roundtables on Competition Policy Papers, No. 311.
[133] OECD (n 124); Lianos (n 130) and OECD (n 132).
[134] OECD (n 124) and OECD (n 132). See also Ohlhausen (n 27).
[135] OECD (n 124).
[136] Heiko Richter, 'The law and policy of government access to private sector data ('B2G data sharing')' in German Federal Ministry of Justice and Consumer Protection, Max Planck Institute for Innovation and Competition (eds), *Data Access, Consumer Interests and Public Welfare* (Nomos, 2021).

mandatory B2G data sharing could help competition authorities to overcome the reluctance of private companies to share their private information, without an official Request for Information, which can be issued only when an investigation has already begun.[137] "Mandatory access rules"[138] need to be data protection and privacy-oriented in order to ensure that personal data are handled in compliance with data protection rules.[139] Furthermore, according to the principle of proportionality that public authorities are bound to respect, a balancing exercise between the need to access privately held data by public authorities and the consequent interference with the fundamental rights of privacy and freedom to conduct a business would have to be carried out.[140]

Rules that impose mandatory access to (specific types of) data can be found in other fields of law, for instance in the financial sector, to grant access to real-time market data;[141] for research and statistical purposes;[142] and in the field of fuel retail market to oblige petrol stations to communicate

[137] Article 18 Regulation (EC) No 1/2003 of 16 December 2002 on the implementation of the rules on competition laid down in Articles 81 and 82 of the Treaty; Richter (n 135); Hofmann and Lorenzoni (n 124). In this regard, it is worth noting that according to the interpretation of the Court of Justice in C-690/20 P *Casino, Guichard-Perrachon SA and AMC v Commission* [2023] EU:C:2023:171, no distinction should be made on information collected before or after the opening of a formal investigation. The case concerned interviews conducted without the necessary safeguards of recording, which led to annul the Commission's decision. According to the Court of Justice "[…] there is nothing in the wording of Article 19(1) of Regulation No 1/2003 or Article 3 of Regulation No 773/2004 to suggest that the application of that recording obligation is contingent on whether the interview conducted by the Commission took place before the formal opening of an investigation in order to collect indicia of an infringement, or afterwards, for the purpose of gathering evidence of an infringement. […] Those provisions do not in any way make the application of the obligation to record contingent on whether the information constituting its subject matter may be categorised as indicia or evidence." [87-88]. Hence, when collecting information outside the Commission's investigatory powers in order to gather *indicia* of an infringement, adequate safeguards that respect fundamental rights of the parties concerned should be adopted.

[138] Richter (n 136).

[139] Ibid. On this matter, it is important to mention that the Court of Justice has given relevance to the right to protect personal data and the right to respect private life, as enshrined in the Charter of Fundamental Rights vis-à-vis the investigatory powers of the Commission. For instance, in Order C-90/24 P(R) *Vivendi SE v Commission* [2024] EU:C:2024:318, the Court stated that the request for information sent to an undertaking was liable to breach the right of privacy of some of the company's employees and that the safeguards provided by the Commission, including the obligation of professional secrecy of Commission's officials, were not sufficient to prevent a breach of the right to respect for private life of the company's employees.

[140] Herwig C H Hofmann, Dirk A Zetzsche and Felix Pflücke, 'The changing nature of 'Regulation by Information': Towards real-time regulation?' (2022) 28 European Law Journal 172. On this matter, see for instance C-470/21 *La Quadrature du Net and Others* [2024] EU:C:2024:370, where the Court of Justice stated that a fair balance must be struck between "on the one hand, the legitimate interests relating to the needs of the investigation […] and, on the other hand, the fundamental rights to respect for private life and protection of personal data of the persons whose data are concerned by the access" [125]. Furthermore, "as can be seen from Article 52(1) of the Charter, that provision allows limitations to be placed on the exercise of those rights, provided that those limitations are provided for by law, that they respect the essence of those rights and that, in compliance with the principle of proportionality, they are necessary and genuinely meet objectives of general interest recognised by the European Union or the need to protect the rights and freedoms of others" [70]. See further on the principle of proportionality C-511/18, C-512/18 and C-520/18 *La Quadrature du Net and Others* [2020] EU:C:2020:791 and C-817/19 *Ligue des droits humains* [2022] EU:C:2022:491. See also C-548/21 *Bezirkshauptmannschaft Landeck* [2024] EU:C:2024:830 where "arising from Article 52(1) of the Charter, […] any limitation on the exercise of a fundamental right must be 'provided for by law', that requirement implying that the legal basis authorising such a limitation must define its scope sufficiently clearly and precisely" [98].

[141] Hofmann, Zetzsche and Pflücke (n 140).

[142] See for instance the UK Digital Economy Act <https://www.legislation.gov.uk/ukpga/2017/30/section/80> accessed on 23 July 2024. See also Commission, 'Towards a European strategy on business-to-government data

changes in the fuel retail price.[143] If examples of this kind that impose B2G data sharing can be found in rather sectorial and sporadic legislations, latest horizonal legislations are unlikely to be used as a legal basis to impose B2G data sharing in competition law. For instance, the Data Act[144] provides a general obligation to make data available to public sector bodies only on the basis of an "exceptional need", which is in case of a public emergency,[145] or in non-emergency situations but only insofar as non-personal data are concerned, to carry out a specific task in the public interest that has been explicitly provided by law, and only if all other means to access such data have been exhausted.[146] Competition authorities are unlikely to satisfy such provisions of non-emergency situations, as Article 16(2) Data Act specifies that B2G data sharing "shall not apply to public sector bodies, the Commission, the European Central Bank or Union bodies carrying out activities for the prevention, investigation, detection or prosecution of criminal or administrative offences or the execution of criminal penalties, or to customs or taxation administration." Data that competition authorities need for screening and market monitoring may fall within their detection activities and therefore outside the scope of this regulation.

Overall, B2G data sharing in competition law could be imposed with a sectorial regulation, along the same line of other sector-specific regulations, rather than relying on a general horizontal legislation. As in the case of B2B data sharing, sectorial regulations may be more tailor made to deal with specific needs and according to the principle of legal certainty, "data requests from public sector bodies […] to data holders should be specific, transparent and proportionate in their scope of content and their granularity [and t]he purpose of the request and the intended use of the data requested should be specific and clearly explained […]".[147]

4.2. Enhancing G2G data sharing

The second mechanism for competition authorities to obtain relevant data in machine readable format is to enhance cooperation and data sharing among public authorities (G2G data sharing).[148] Within the framework of the EU data strategy, which aims to create a "Common European Data Space",[149] mechanisms for ensuring the free flow of data and cooperation among public authorities are laid down, where "sharing should become a default".[150] Firstly, cooperation among competition authorities should not be limited to share best practices and information on ongoing investigations,

---

sharing for the public interest' (Final report prepared by the High-Level Expert Group on Business-to-Government Data Sharing 2020) 35 <https://op.europa.eu/en/publication-detail/-/publication/d96edc29-70fd-11eb-9ac9-01aa75ed71a1> accessed 27 August 2024 and Richter (n 136).

[143] See for instance <https://www.bmwk.de/Redaktion/EN/Artikel/Energy/market-transparence-units.html> accessed 27 August 2024 and Richter (n 136).

[144] Regulation (EU) 2023/2854 of 13 December 2023 on harmonised rules on fair access to and use of data and amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828 (Data Act).

[145] Article 15(1)(a) Data Act.

[146] Article 15(1)(b) and Recital 65 Data Act.

[147] Recital 69 Data Act.

[148] See OECD (n 132) according to which "co-operation between competition authorities and other domestic entities (such as government bodies, procurement agencies and sector regulators) may facilitate the acquisition of data and is therefore crucial for achieving effective screening tools" 10.

[149] <https://digital-strategy.ec.europa.eu/en/policies/data-spaces> accessed on 24 July 2024.

[150] Regulation (EU) 2024/903 of 13 March 2024 laying down measures for a high level of public sector interoperability across the Union (Interoperable Europe Act) Recitals 25 and 26.

but *ex ante* data sharing should occur already at the detection phase. AI systems developed for screening purposes require a large dataset. "Does such a data set exist today – […] with a sufficient number of cases of both collusion and not-collusion, with the necessary data on price, cost, and drivers of supply and demand […]?"[151] As it might be almost impossible to manually create such a database, data on past cartel cases should be shared among competition authorities to create a large data pool that can be accessed whenever algorithms need to be trained.[152] Being able to dispose of such dataset would have the advantage of providing competition authorities with enough data (*quantitative data*) of relevant information of past cases (*qualitative data*) in *machine readable format* that can be used to feed their algorithms, enhancing the performance of their AI tools, in terms of accuracy (with less false positives and false negatives) and reliability of the results. It could even be suggested that if digital screening tools are trained on a dataset which is non-biased, accurate and sufficiently large, their outcomes could be used as direct evidence to issue inspection decisions and prove an anticompetitive behaviours, under supervision of a human being.[153] This approach would significantly increase competition authorities' efficiency and detection capabilities.

As part of G2G data sharing, cooperation between competition authorities and other public enforcers should be encouraged.[154] For example, cooperation with public procurement bodies is often needed for competition authorities to have access to data concerning tenders. Competition authorities often investigate bid rigging cases, and in order to use accurate digital tools embedded with machine learning and AI techniques, they need to have access to structured bidding data that usually only public procurement bodies possess.

Enhancing data sharing between competition authorities among each other and between them and public procurement bodies requires "good working relationships" and "a high degree of trust between authorities".[155] Latest EU digital legislations also pinpoint the need to create a trust environment where data can freely move within the single market as a fifth freedom, where rights and interests of data holders and data subjects are respected.[156] In particular, when personal data are the subject of G2G data sharing, compliance with data protection law can be ensured through anonymisation and pseudonymization techniques, as suggested for B2B data sharing. Finally, also in the context of G2G data sharing, a data trustee can be a solution as it may be regarded as a trustworthy intermediary body which supervises that data sharing among public authorities occur in respect of all necessary safeguards for data subjects and data holders.

---

[151] Rosa M Abrantes-Metz and Albert D Metz, 'Can Machine Learning aid in Cartel Detection?' (2018) CPI Antitrust Chronicle 1.
[152] Lianos (n 130) 16 and OECD (n 124).
[153] OECD (n 124).
[154] OECD (n 132) "[…] co-operation between competition authorities and other domestic entities (such as government bodies, procurement agencies and sector regulators) may facilitate the acquisition of data and is therefore crucial for achieving effective screening tools" 10.
[155] OECD (n 124) 26.
[156] Commission, 'Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. A European strategy for data' COM(2020) 66 final; Borgogno and Colangelo (n 57).

5. Conclusions

This paper has analysed how data can impact competition law, from both sides, market players and competition authorities. On the one hand, modern business models revolve around the collection and processing of personal and non-personal data to generate income and deliver new products, as innovative technologies depend on data. In a digital and data-driven economy, the risk is that few large companies can use data advantage to strengthen and abuse their dominant position at the expenses of smaller competitors. Therefore, solutions to impose data sharing among businesses (B2B data sharing) have been laid down. First, competition tools can play a role to impose *ex post* data access according to the essential facility doctrine, which however has been criticized when applied to data, and some adaptations may be required. Second, *ex ante* rules that impose B2B data sharing can be implemented in the form of sectorial and horizontal legislations. They provide compulsory B2B data sharing rules to specific data holders to remedy to market failures, before affecting competition, ensuring a level playing field in data-centric markets.

On the other hand, besides focusing on how data has impacted competition law from the point of view of market players, this paper has showed how data has become an essential factor for competition authorities when enforcing competition law. More and more digital tools have started to be developed by enforcers for which data availability is of utmost importance. Hence, data sharing between competition authorities and private businesses (B2G data sharing) and between them and other public authorities (G2G data sharing) may be a long-term solution to ensure enforcers to have access to qualitative and quantitative data in machine readable format, also in the light of the latest EU digital legislation. In a data-driven economy, data is a key element for both private and public actors, and B2G and G2G data sharing for the purpose of enforcing competition law "should become a default".[157]

[157] Recitals 25 and 26 Interoperable Europe Act.